

Clasificación de medidas de glucemia en función de ingestas en diabetes gestacional

<E>. Caballero Ruiz¹, <G>. García-Saez¹, <M>. Rigla Cros², <M>. Balsells³, . Pons², <EJ>. Gómez Aguilera¹, <ME>. Hernando Pérez¹

¹ Grupo de Bioingeniería y Telemedicina, Universidad Politécnica de Madrid, España, {ecaballero, ggarcia, egomez, elena}@gbt.tfo.upm.es

² Servicio de Endocrinología y Nutrición, Hospital de Sabadell, Barcelona, España, mrigla@tauli.cat, bpons2@gmail.com

³ Servicio de Endocrinología y Nutrición, Hospital Mutua de Terrassa, Barcelona, España

Resumen

Este trabajo presenta un clasificador de medidas de glucemia en función de las ingestas asociadas para pacientes con diabetes gestacional. Se presentan los resultados obtenidos al comparar la relevancia de diferentes atributos así como del uso de dos de los algoritmos más populares en el mundo del aprendizaje automático: las redes neuronales y los árboles de decisión. El estudio se ha realizado con los datos de 53 pacientes pertenecientes al Hospital de Sabadell y al Hospital Mutua de Terrassa obteniendo un 91,72% de precisión en el caso de la red neuronal, y un 95,92% con el árbol de decisión. La clasificación automática de medidas de glucemia permitirá a los especialistas pautar un tratamiento más acertado en base a la información obtenida directamente del glucómetro de las pacientes, contribuyendo así al desarrollo de los sistemas automáticos de ayuda a la decisión para diabetes gestacional.

1. Introducción

La diabetes gestacional (GDM) se define como una intolerancia a los carbohidratos que provoca una hiperglucemia de gravedad variable y que comienza o se diagnostica por primera vez durante el embarazo. Este elevado nivel de glucosa en sangre se transfiere al feto [1] provocándole diversos trastornos: crecimiento excesivo de los tejidos adiposos, lo que aumenta la necesidad de cesárea, hipoglucemia neonatal y aumento del riesgo de muerte fetal intrauterina. Además aumenta el riesgo de padecer diabetes tipo 2 una vez finalizado el periodo de gestación tanto para la madre como para el feto.

En el proyecto SINEDiE se propone el desarrollo de herramientas inteligentes y de educación para la GDM basadas en técnicas neurodifusas integradas en un sistema de telemedicina. Los sistemas de telemedicina han sido utilizados con éxito en numerosas ocasiones en diabetes [2,3] y la integración de herramientas de ayuda a la decisión en este tipo de sistemas ayuda a una mejor interpretación de los datos [4].

En GDM la glucemia se regula mediante la dieta, el ejercicio físico y en algunas ocasiones es precisa la administración de insulina. Para ello las pacientes deben

controlar sus niveles de glucosa en sangre con ayuda de un glucómetro, realizando al menos 4 medidas al día, una en ayunas y las otras tres después de cada ingesta principal. Estos datos, nivel de glucosa e ingesta asociada, son apuntados por las pacientes en un libro de control que enseñan semanalmente al especialista. El especialista utiliza esta información para pautar la dieta correspondiente y la administración de insulina si fuera necesario.

La mayoría de los glucómetros permiten almacenar las medidas realizadas por las pacientes, evitando así la tediosa tarea de apuntar los valores en el libro de control. Esto presenta una gran ventaja frente a la posible ausencia y no fiabilidad de los datos apuntados de manera manual, y abre la posibilidad de procesar los datos de manera automática e integrarlos en sistemas de telemedicina y ayuda a la decisión. Los glucómetros almacenan tres datos por cada muestra de sangre analizada: fecha, hora y valor de glucosa en mg/dL. Estos tres datos formarán una medida. Sin embargo, la información obtenida del glucómetro es insuficiente para ser examinada directamente por un especialista, ya que para ellos es esencial conocer la ingesta a la que está asociada cada medida y si se trata de datos preprandiales o posprandiales. Ante los datos almacenados en el archivo de memoria del glucómetro, el especialista se ve obligado a descifrar a qué ingesta corresponde cada medida. Por lo tanto, la lista de valores proporcionados por el glucómetro debe ser clasificada en función de las ingestas para que el especialista pueda determinar el tratamiento correspondiente.

Según la definición de Tom Mitchell [5], se dice que un programa aprende de la experiencia si mejora su rendimiento en la realización de una tarea concreta al aumentar dicha experiencia. Los algoritmos de aprendizaje automático permiten diseñar programas capaces de aprender de datos pasados. En nuestro caso, nos permitirá diseñar un clasificador capaz de aprender de datos de pacientes previamente clasificados, para poder clasificar nuevos datos.

Este trabajo presenta un clasificador de medidas de glucemia para pacientes con GDM comparando dos de los algoritmos más populares de aprendizaje automático supervisado: la red neuronal y el árbol de decisión, además de comparar la utilización de diferente número de atributos de entrada hasta llegar a la solución óptima.

2. Material y métodos

2.1. Metodología de análisis

Se han estudiado diez clasificadores utilizando dos algoritmos de aprendizaje automático diferentes, uno basado en el árbol de decisión C4.5 de Quinlan [6] y otro en una red neuronal con arquitectura Perceptrón Multicapa [7,8].

El árbol de decisión C4.5, presenta una buena precisión para una amplia variedad de problemas además de permitir representar la información de manera comprensible para el personal sin conocimientos específicos. Sin embargo, la desventaja que presentan es que son poco robustos ya que su precisión depende en gran medida de los umbrales fijados en su configuración inicial [9].

Por otro lado tenemos las redes neuronales que presentan una menor precisión frente a los árboles de decisión pero que son más robustas. La desventaja que presenta este método de clasificación es que resulta como un sistema de caja negra sin unos conocimientos específicos del comportamiento, ya que se conocen las entradas y las salidas pero no la función de salida. La arquitectura elegida ha sido la de Perceptrón Multicapa [7,8], con 3 capas, la de entrada, la oculta y la de salida. Al utilizar diferente número de atributos, las neuronas de la capa de entrada variarán de acuerdo a ellos.

Con el objetivo de determinar el modelo óptimo se han probado diferente número de atributos ya que mientras los árboles de decisión parecen comportarse mejor con un mayor número de entradas, las redes neuronales no siguen este comportamiento.

2.2. Datos de glucemia

Se dispone de un total de 9175 datos de un total de 53 pacientes proporcionados por el Hospital de Sabadell y el Hospital Mutua de Terrassa. Las pacientes fueron provistas de un glucómetro con el cual se midieron la glucemia entre 3 y 6 veces al día durante un periodo que comprendía desde que se realizaba el diagnóstico de GDM hasta el momento del parto. Las pacientes también anotaban las medidas en sus libros de control. Los datos almacenados en los glucómetros fueron clasificados en función de las ingestas por un experto, sin tener en cuenta las anotaciones de las pacientes.

2.3. Selección de parámetros de entrada y salida

Del archivo de memoria del glucómetro y de la historia clínica de la paciente obtenemos directamente 4 parámetros de entrada ("*abs_date*", "*abs_time*", "*bg*" e "*insulin*"). Éste último atributo indicará si la paciente tiene activo un tratamiento con insulina y de qué tipo. La insulina afecta al valor de la glucemia posprandial. A

partir de ellos calcularemos 6 atributos más para poder evaluar el comportamiento de los clasificadores al incrementar su complejidad. Este incremento se conseguirá utilizando un número mayor de atributos de entrada.

Los hábitos horarios y alimenticios de las pacientes pueden variar según se trate de un día laborable o un fin de semana, por lo tanto parece lógico pensar que podría afectar a la clasificación y por ello se ha calculado el día de la semana al que pertenece cada medida. Se obtiene así el nuevo atributo de entrada: "*dow*".

Los tiempos transcurridos entre medidas también son relevantes, ya que a partir de ellos se puede saber, por ejemplo, si se trata de la primera medida del día, o si se trata de una medida repetida. Se han calculado dos atributos para proporcionar esta información: "*interval_prev*", tiempo transcurrido desde la anterior medida, e "*interval_post*", tiempo transcurrido hasta la siguiente medida.

Contar con los niveles de glucemia de medidas consecutivas nos puede proporcionar información sobre si ha habido una ingesta de comida entre ellas. Por lo tanto, se han calculado otros dos atributos: "*bg_prev*", valor de glucemia de la anterior medida y "*bg_post*", valor de glucemia de la medida posterior.

El último atributo de entrada calculado será el "*intake*". Este atributo representa la ingesta más probable según los hábitos de la paciente y podrá tener los valores desayuno, comida o cena. Para obtenerlo, se realiza primero un clustering para agrupar las medidas de cada paciente en 3 subgrupos en función de los atributos "*bg*" y "*abs_time*". Para ello se ha implementado el algoritmo de las k-medias [10] en Octave [11]. A partir de los 3 subgrupos obtenidos, extraeremos los rangos horarios de la paciente examinando el mínimo y máximo valor de "*abs_time*" de cada grupo. Con estos rangos horarios, el subgrupo al que pertenece cada medida, los atributos calculados anteriormente y un conjunto de reglas lógicas obtendremos finalmente el atributo "*intake*".

Se utilizarán diez clases diferentes para realizar la clasificación de las medidas: "*break-prep*", "*break-post*", "*lunch-prep*", "*lunch-post*", "*dinner-prep*", "*dinner-post*", "*morning*", "*afternoon*", "*night*" y "*repeated*". Las seis primeras corresponden al desayuno, comida y cena preprandiales y posprandiales. Las tres siguientes corresponden a la mañana, tarde y noche y la última se utiliza para medidas repetidas, ya que las pacientes se repiten las medidas cuando no se sienten bien o no están satisfechas con el resultado obtenido.

2.4. Clasificadores

Para cada algoritmo, el árbol C4.5 y la red neuronal, se han creado cinco clasificadores con diferente número de atributos:

- *3 atributos*: Valores procedentes directamente del glucómetro: "*abs_date*", "*abs_time*" y "*bg*".
- *4 atributos*: Definidos por un experto en un estudio anterior con pacientes con diabetes tipo 1 [12]. Los

atributos utilizados fueron: “abs_time”, “bg”, “interval_prev” e “insulin”.

- 8 atributos: Todos menos el atributo “intake” y el “insulin”. Ambos atributos son los únicos que presentan un coste no despreciable en su cálculo y por lo tanto es interesante estudiar su repercusión en la precisión del clasificador.
- 9 atributos: Prescindimos únicamente del atributo “intake”, ya que puede que no dispongamos de suficiente información para obtenerlo o simplemente resulte demasiado costoso realizar el clustering previo de los datos de cada paciente.
- 10 atributos. Todos los atributos disponibles.

2.5. Metodología de evaluación

Los diferentes modelos se ejecutarán en un Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz utilizando la herramienta Weka [13], ya que dispone de la implementación de los dos algoritmos seleccionados, el MLP para la red neuronal y el J48 para el árbol C4.5.

A la hora de evaluar la precisión de los clasificadores, se utilizará el método de validación cruzada de k iteraciones [14]. La validación cruzada consiste en dividir los datos en n partes, y por cada parte, se entrena el clasificador con las n – 1 partes restantes y se prueba con la parte seleccionada. Así por cada una de las n particiones. Este método permite evaluar el comportamiento del clasificador frente a datos diferentes a los utilizados para su entrenamiento. También se evaluará el clasificador utilizando todos los datos disponibles tanto para su entrenamiento como para la evaluación.

3. Resultados

En la tabla 1 se pueden ver los resultados obtenidos tras las ejecuciones en Weka de las distintas combinaciones de algoritmos y números de atributos de entrada. En la tercera columna “Cjto. Entr.” se muestra el resultado de entrenar el clasificador con todos los datos disponibles y luego utilizar el mismo conjunto de datos para evaluar la precisión. Sin embargo, nosotros nos fijaremos en la cuarta columna, donde se muestra el resultado obtenido mediante el método de validación cruzada, ya que lo que nos interesa es estudiar el comportamiento de los clasificadores frente a datos nuevos.

	Algoritmo	Conj. Entr.	Validación Cruzada
10 Atributos	MLP	93,82%	91,72%
	J48	98,01%	95,92%
9 Atributos	MLP	91,78%	89,57%
	J48	97,84%	95,54%
8 Atributos	MLP	90,21%	89,56%
	J48	98,43%	95,53%
4 Atributos	MLP	89,53%	88,68%
	J48	96,41%	94,30%
3 Atributos	MLP	86,65%	84,53%
	J48	90,54%	87,48%

Tabla 1. Precisión de los diferentes modelos

Con ambos algoritmos se ha logrado una precisión de más del 90%, 91,72% con el MLP y 95,51% con el J48, aunque se observa que ésta se reduce a medida que disminuimos el número de atributos de entrada.

Para hacer un análisis más exhaustivo y poder determinar en que casos se comete mayor número de fallos, se examina la matriz de confusión de la ejecución más satisfactoria de ambos algoritmos, es decir, con diez atributos de entrada y la técnica de validación cruzada que es la que nos interesa evaluar.

Las clases en las que ambos algoritmos cometen más fallos son: “afternoon”, “morning”, “repeated”, “night”, “lunch-prep” y “dinner-prep”. Esto puede ser debido a que son las clases minoritarias en nuestro conjunto de datos. Concretamente solo el 13,45% de los datos proporcionados por los hospitales pertenecen a estas clases. Aparecen menos frecuentemente porque las pacientes no suelen realizarse este tipo de medidas, ya que los especialistas sólo les recomiendan que hagan las otras cuatro. Los clasificadores se comportan mucho mejor con las clases de las que tenemos más datos de entrenamiento: “break-prep”, “dinner-post”, “break-post” y “lunch-post”.

Otro aspecto importante a tener en cuenta es el tiempo de construcción de los modelos, dato que se muestra en la tabla 2. En el caso de la red neuronal, se muestra el tiempo de una iteración en la validación cruzada.

Tiempo (seg)	10 Atrib.	9 Atrib.	8 Atrib.	4 Atrib.	3 Atrib.
MLP	58,66	48,15	23,85	35,7	15,84
J48	0,23	0,13	0,13	0,07	0,08

Tabla 2. Tiempo de construcción de los diferentes modelos

El tiempo de construcción del modelo basado en la red neuronal es mucho mayor que el basado en el árbol de decisión. Además este tiempo aumenta en ambos algoritmos conforme vamos añadiendo atributos debido al aumento de la complejidad de los modelos. Sin embargo, esto no ocurre en la ejecución con 4 y 8 atributos, donde al incluir 4 atributos más, el tiempo de ejecución disminuye. Otro aspecto destacable es en el caso de la utilización de 8 y 9 atributos, donde al incluir el atributo “insulin” el tiempo de construcción de la red neuronal aumenta en un 50% sin un aumento de precisión relevante.

4. Discusión

Tanto con los modelos basados en el árbol de decisión C4.5, como en los basados en el Perceptrón Multicapa se han conseguido resultados muy satisfactorios, 95,92% y 91.72% de precisión respectivamente.

A la vista de los resultados obtenidos, destacaremos la notable mejora en la precisión de los clasificadores al añadir, a los tres datos proporcionados por el glucómetro, un cuarto atributo relativo a la insulina. Es evidente que la mejor configuración es la de mayor número de entradas, sin embargo hay que tener en cuenta que la obtención de los atributos “intake” e “insulin” es más compleja que la de los demás.

La realización del clustering para la obtención del atributo “intake” presenta una ventaja frente a su obtención mediante un rango fijo de horarios, ya que cada paciente tiene unos hábitos diferentes. Sin embargo habría que estudiar el coste adicional que supone su obtención, teniendo en cuenta que la precisión sin este atributo sólo disminuye en un 2% en el caso de la red neuronal y de apenas un 0,38% en el caso del árbol de decisión. Aunque menos fiable, también se podría obtener este atributo preguntando a cada paciente sus horarios habituales de ingesta de cada comida.

Otro aspecto a tener en cuenta es que la precisión de la ejecución con 8 y 9 atributos es prácticamente la misma, 89,56% y 89,57% en el caso de la red neuronal y 95,92% y 95,54% en el caso del árbol de decisión. Por lo tanto, si se decidiera prescindir del atributo “intake” y seleccionar un clasificador con los otros 9 atributos, también deberíamos prescindir del atributo “insulin” e implementar el clasificador con 8 atributos de entrada. De esta manera, se simplifica el modelo y se disminuye el coste de construcción a la mitad en el caso de la red neuronal.

Se observa que los modelos basados en el árbol de decisión obtienen mejores resultados que los basados en la red neuronal en todos los casos. Sin embargo, está técnica ha recibido numerosas críticas debido a su poca robustez, por lo tanto, deberíamos seguir estudiando su comportamiento antes de efectuar una selección definitiva del algoritmo a utilizar para diseñar nuestro clasificador. Además la precisión obtenida con la red neuronal es tan solo un 4% menor que la obtenida con el árbol de decisión, por lo tanto la seguiremos teniendo en cuenta en futuros estudios. También se compararán en un futuro las anotaciones de los libros de control de las pacientes con la clasificación hecha por el experto y la obtenida con nuestro clasificador.

5. Conclusiones

Tras examinar los resultados se concluye que es posible determinar las ingestas asociadas a las medidas de glucemia de pacientes con GDM con una precisión del 95,92% en el caso del árbol de decisión y 91,72% en el caso de la red neuronal. Un análisis más exhaustivo del diseño de la red neuronal podría aumentar la precisión obtenida con este algoritmo, ya que se ha observado que un aumento de complejidad repercute positivamente en los resultados obtenidos. Además, la mayoría de errores se cometen en medidas poco ejemplificadas, por lo que conseguir más datos de entrenamiento de esas clases podría aumentar la precisión del clasificador.

La clasificación automática de las medidas de glucemia permitirá mostrar a los especialistas la información obtenida directamente del glucómetro, la cual es más completa que la apuntada manualmente por las pacientes. Contar con una información más precisa ayudará al especialista a pautar un tratamiento más acertado y evitar así los trastornos asociados a la diabetes gestacional. Además, el tratamiento digital de los datos es esencial para los sistemas de telemedicina, cuya eficacia aumentará indudablemente al disponer de datos más ricos.

Agradecimientos

Este trabajo ha sido financiado por el proyecto del Fondo de Investigaciones Sanitarias ‘SINEDI’: Sistemas INteligentes y de Educación para el control de la Diabetes diagnosticada en el Embarazo (PI10/01125). Agradecer al Hospital de Sabadell y al Hospital Mutua de Terrassa su colaboración.

Referencias

- [1] National Collaborating Centre for Women's and Children's Health. Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. Clinical guideline. RCOG Press, 2008 (ISBN: 978-1-904752-47-9).
- [2] Verhoeven F. Gemert-Pijnen L. Dijkstra K. Nijland N. Seydel E. Stehouder M. The Contribution of Teleconsultation and Videoconferencing to Diabetes Care: A Systematic Literature Review. *J Med Internet Res*, vol. 9, Issue 5, 2007: e37.
- [3] Verhoeven V. Tanja-Dijkstra K. Nijland N. Eysenbach G. Gemert-Pijnen L. Asynchronous and Synchronous Teleconsultation for Diabetes Care: A Systematic Literature Review, *J Diabetes Sci Technol*, vol. 4, Issue 3, 2010, 666-684.
- [4] Hernando M, Gómez E, Corcoy R y del Pozo F. Evaluation of DIABNET, a decision support system for therapy planning in gestational diabetes, *Comp. Meth. and Prog. in Biomedicine*, vol. 62, Issue 3, 2000, pp. 235-248.
- [5] Mitchell TM. Machine Learning. McGraw-Hill, 1997. (ISBN: 0-07-042807-7).
- [6] Quinlan JR. Induction of decision trees. *Machine Learning*, vol 1, num, 1, 1998, pp 81-106 (ISSN: 0885-6125).
- [7] Anderson JA. An introduction to neural networks. MIT Press, 1995. (ISBN: 0-262-03224-4).
- [8] Haykin S. Neural networks: a comprehensive foundation, Prentice Hall, 1998.
- [9] Quinlan JR. Induction of decision trees. *Machine Learning*, vol. 1, num. 1, 1998, pp. 81-106 (ISSN: 0885-6125).
- [10] Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* Vol. 28, No. 1, 1979, pp. 100-108.
- [11] Página web de GNU Octave. <http://gnu.org/software/octave> (Consultada: Septiembre 2012)
- [12] García-Sáez, G, Alonso JM, Molero J, Rigla M, Martínez-Sarriegui I, de Leiva, A, Gómez E, Hernando, M. Mealtime Blood Glucose Classifier Based on Fuzzy Logic for the DIABTel Telemedicine System. *Artificial Intelligence in Medicine, Lecture Notes in Computer Science*, Volume 5651, 2009, pp 295-304 (ISSN: 978-3-642-02975-2)
- [13] Witten H, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, 1999 (ISBN: 978-0-12-374856-0).
- [14] Sakata SW. Cross-validation estimates IMSE. *Advances in Neural Information Processing Systems 6th*. 1994, pp. 391-398.